



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Cormack-Jolly-Seber models: time and age perspectives

**Citation for published version:**

Sarzo, B, Conesa, D & King, R 2020, 'Cormack-Jolly-Seber models: time and age perspectives', *Stochastic Environmental Research and Risk Assessment*, vol. 34, pp. 1683–1698. <https://doi.org/10.1007/s00477-020-01840-x>

**Digital Object Identifier (DOI):**

[10.1007/s00477-020-01840-x](https://doi.org/10.1007/s00477-020-01840-x)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Stochastic Environmental Research and Risk Assessment

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Cormack-Jolly-Seber models: time and age perspectives

Blanca Sarzo\* · David Conesa · Ruth King

Received: date / Accepted: date

**Abstract** Survival is a key demographic characteristic in many areas including both human demography and population ecology. However, it is often the case that data collection protocols are different in these areas, resulting in different models and methods of analysis. This paper is motivated for the different emphasis given to the elicitation of the temporal scale (and consequently, on the origin time) in ecological and medical survival studies. Specifically, in medical studies, the origin time is often determined in advance with individuals followed over a period of time at regular (or irregular) intervals, thus focusing on time within study (or age to a given reference point). However, in ecological capture-recapture studies, the capture occasions are typically fixed in advance, with an imperfect detection process observing individuals at these times. Moreover, the temporal scale is often primarily specified at the capture occasion level. In this work we focus on an ecological capture-recapture study related to guillemots and compare and contrast two different temporal scales: (i) calendar (or capture occasion); and (ii) age (or time within study), in terms of the way the data may be represented and in relation to the ecological Cormack-Jolly-Seber-type model. The different temporal scales provides insights into the different underlying structures, which can then be combined into a joint (calendar and age) dependence model.

**Keywords** Age scale · Bayesian inference · Calendar scale · Capture-recapture · Temporal scale · Survival analysis

## 1 Introduction

Survival analysis is an area of research of special relevance in many fields such as agronomy, ecology, engineering and medicine (among others). Although in all these fields the final objective is often the same, consisting of the estimation of time to the occurrence of events,

---

Blanca Sarzo  
Department of Statistics and O.R., University of Valencia, Spain  
E-mail: Blanca.Sarzo@uv.es

David Conesa  
Department of Statistics and O.R., University of Valencia, Spain

Ruth King  
School of Mathematics, University of Edinburgh, United Kingdom

differences on the way data are collected often result in different models and associated statistical methods. We discuss the different ideas associated within the application areas of medicine and ecology, and in particular focus on the natural temporal scales that arise from the different data collection processes, with the aim of borrowing ideas from medicine to better understand the representation and modelling of ecological capture-recapture data.

In the context of medical survival studies, one of the main requirements which is commonly performed at the beginning of the experiment, is that the origin time (time zero) must be unambiguously defined for each individual. The time scale of the study also needs to be specified. See Table 1 for different choices of origin time and time scales. Indeed, the use of different temporal scales may result in different inferences and interpretations (28; 29; 8; 49).

Table 1: Examples of possible choices of origin time and time scales in medical studies extracted from (20).

Time zero	Time scale
Birth	Age
Diagnosis of disease	Duration
Entry into state	Waiting time
Bleeding	Duration of pregnancy
Start of treatment	Length of treatment
Baseline measurement	Calendar time

In ecological studies, survival probabilities (and hence time to death) are often the main interest and monitoring data are frequently obtained using a capture-recapture protocol (although there are other data collection methods). This involves a series of capture occasions whereby at each occasion individuals are observed, uniquely identified (possibly by artificial means such as tag or ring) and released. For simplicity, we assume a single capture occasion each calendar year. The data correspond to the capture histories of each individual observed within the study. A large class of associated capture-recapture models have been developed over many decades to estimate the survival probabilities of interest (31; 42; 27; 26; 47). In particular, the Cormack-Jolly-Seber (CJS) model (11; 21; 46) is the basis for many open population capture-recapture-type models, including, for example, capture-recapture-recovery models (with the inclusion of additional dead recoveries; 5; 6; 44), multi-state (Arnason-Schwarz) models (for discrete time-varying individual states; 4; 45; 23; 25), multi-event models (that allow for additional state-uncertainty; 39), continuous time-varying individual covariate models (1; 24; 30); spatially explicit capture-recapture models (including additional spatial information; 2); random effect models (42); and stopover models (modelling recruitment into the population; 37).

In the traditional CJS-type models, the (*apparent*) survival probabilities (referred as *apparent* since death and unavailable for recapture are usually confounded) are estimated and expressed in terms of annual survival (or survival between capture occasions) due to, in general, wild animals having a natural yearly life-cycle. However, more importantly, the elicitation of the temporal scale is not always addressed in such studies, the standard choice often being the capture-occasion (or calendar) scale. In this context, for standard CJS models, the time zero for a given observed individual naturally corresponds to the capture-occasion on which the individual is marked (and hence enters into the study at this time), and the model conditions on this initial capture time for the given individual. This can be extended when

age at capture is recorded, leading to time zero corresponding to “age 0” (which may be prior to the initial capture occasion), and the development of age-dependent models which often include a “first-year” or “immature” dependence to account for young individuals typically having a higher mortality rate than older birds (34; 33). Stopover models further extend these ideas, by treating time zero as unknown within the model so that the time of arrival in the study may be before their first capture, also extending the set of parameters to include arrival probabilities, and removing the conditioning on the initial capture time for each individual.

In this work we present a comparison between two different temporal scales (age and calendar) and discuss the presentation of the data using the different scales in the context of ecological capture-recapture models. We illustrate the usefulness of age scale in this context (typically used in medical studies) as well as describe both the differences and similarities in the notation and interpretation of parameters depending on the scale used. In particular, we present the CJS model in the alternative age scale (both in state-space and multinomial model formulation) before incorporating both age and time dependence. We apply the models to the largest colony of common guillemot (*Uria aalge*) in the Baltic Sea (Stora Karlsö, Sweden) to estimate the juvenile survival probability of interest.

After this introduction, the remaining of the paper is as follows. Section 2 presents the notation associated with capture-recapture studies and the different parameterisations depending on the scales, while in Section 3 we extend the dependence structure to age and time and discuss the general expression of the CJS models in both state-space and multinomial formulations. Section 4 presents the application of the model in a real example, and finally Section 5 concludes.

## 2 Notation

In this section we present the notation associated with capture-recapture studies, and in particular focus on the different parameterisations based on the different temporal scales of the model parameters (i.e. capture occasion or age). Further we discuss the associated representation of the data for the different temporal scales.

Let  $i = 1, \dots, N$  denote the individuals observed within the study period; and  $t = 1, \dots, T$  the associated capture occasions. For notational simplicity we assume that the capture occasions correspond to an occasion within a calendar year, so that we refer to the temporal scale of the capture occasion to be (calendar) year (but this is clearly more general). Further, we define  $\mathcal{A} = \{0, \dots, A\}$  to be the set of possible (natural) ages for an individual within the study period (age 0 corresponds to individuals in their first year of life). Note that in practice (as for our case study) the upper limit,  $A$ , will often correspond to an individual of at least age  $A$  (for example, when it becomes an adult). We let  $a_{i0}$  denote the age of individual  $i$  at initial capture; and  $\mathbf{a}_0 = \{a_{i0} : i = 1, \dots, N\}$  the set of initial ages for the individuals observed within the study. Further we let  $\mathcal{A}_0 = \{0, \dots, A_0\}$  denote the set of initial ages. If  $a_{i0} = 0$  for all  $i = 1, \dots, N$  (i.e. all individuals enter into the study at age 0), then their age at each subsequent capture occasion is equal to the length of time an individual would have been in the study, akin to the temporal scale of medical studies of time (so that  $A_0 = 0$  and  $A = T - 1$ ). Alternatively, if individuals may be observed at different ages at initial capture, the combination of their initial age at capture and time since initial capture corresponds to their (natural) age at any given capture occasion (and the maximum age of an individual within the study is equal to  $A = A_0 + T - 1$ ).

## 2.1 Data representation

The data correspond to the set of observed capture histories of each individual observed within the study. This is typically expressed in terms of the capture occasion as the temporal scale. We let:

$$y_{it} = \begin{cases} 0 & \text{individual } i \text{ is not observed at capture occasion } t; \\ 1 & \text{individual } i \text{ is observed at capture occasion } t. \end{cases}$$

The capture history of individual  $i$  is then denoted  $\mathbf{y}_i = \{y_{it} : t = 1, \dots, T\}$ ; and full set of capture histories by  $\mathbf{y} = \{\mathbf{y}_i : i = 1, \dots, N\}$ . If individuals can be observed at different ages, this is also recorded as an observed (discrete) covariate for each individual. Further we let  $f_i$  denote the capture occasion on which individual  $i$  is first observed; and set  $\mathbf{f} = \{f_1, \dots, f_N\}$ . Finally we introduce the idea of cohort  $w = 1, \dots, T - 1$  such that individual  $i$  belongs to cohort  $w$  if  $f_i = w$  (i.e. individual  $i$  is observed for the first time on capture occasion  $w$ ). Thus, [although other options are possible, in this work](#) cohort is defined on the capture occasion (or calendar year) scale. We consider the general case where individuals may enter into the study at different ages so that individuals belonging to the same cohort may be of different ages; if all individuals are observed at the same age at initial capture, then cohort and age are interchangeable.

Alternatively, and akin to how such studies are recorded within medical studies, we can present the data in terms of the age of an individual (which in turn can be sub-divided into the initial age of an individual combined with length of time in the study). This represents a focus on the (natural) age of an individual as a primary factor within the study, for example, due to the life cycle of the given species. We note that the transforming of the data to the age scale, if this is the relevant temporal scale within the study, may provide a more useful representation. For example, [the capture histories can be equivalently presented in terms of the capture histories from only the initial capture](#), corresponding to  $\mathbf{y}_i = \{y_{it} : t = f_i, \dots, T : i = 1, \dots, N\}$  (thus removing the trailing zeroes), combined with the initial age  $a_{i0}$ . The corresponding capture histories will then be of different lengths, dependent on time of first capture. [If all individuals are observed at initial age 0, this presentation of the data highlights, for example, the decreasing sample size of individuals as age increases.](#) The difference in representations between the calendar year and age temporal scales of the capture histories is illustrated in the toy example presented in Figures 1 and 2. Figure 1 on the left shows capture-histories in the standard calendar scale, and on the right the same individual histories but in age scale. In order to highlight the differences between the temporal scales, in this simple case we assume that all individuals are age 0 at their initial capture (i.e.  $a_{i0} = 0$  for all  $i$ ). If individuals differ in (natural) age at initial capture then there will be additional “steps” observed in the right hand plot of Figure 1 with age and “time in study” not equivalent to each other. Figure 2 shows cohorts 1, 2 and 3 in age (top) and calendar scale (bottom) corresponding to the same individuals in Figure 1. Note that the later the cohort, the smaller the number of occasions or the maximum “time in study” value (in calendar and age scale, respectively). This can be explained since as the cohort number increases, the number of possible capture occasions after initial capture decreases.

## 2.2 Parameter dependence

The commonly used Cormack-Jolly-Seber model has two sets of parameters corresponding to the survival probabilities and observation (or capture) probabilities. These parameters are

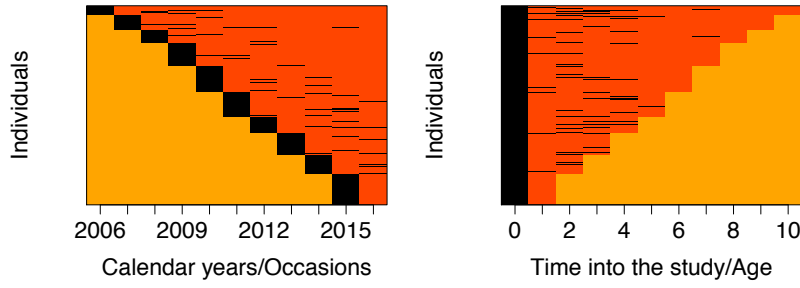


Fig. 1: Toy example representing the different structure of the same individual capture-recapture histories in calendar (on the left) and age scale (on the right) assuming all individuals are zero years old at the beginning of the study (i.e.  $a_{j0} = 0$ ). Rows represent individuals ( $n=300$ ), orange lines represent non information, red lines indicate that individual has not been seen, and black lines represent that the individual has been seen.

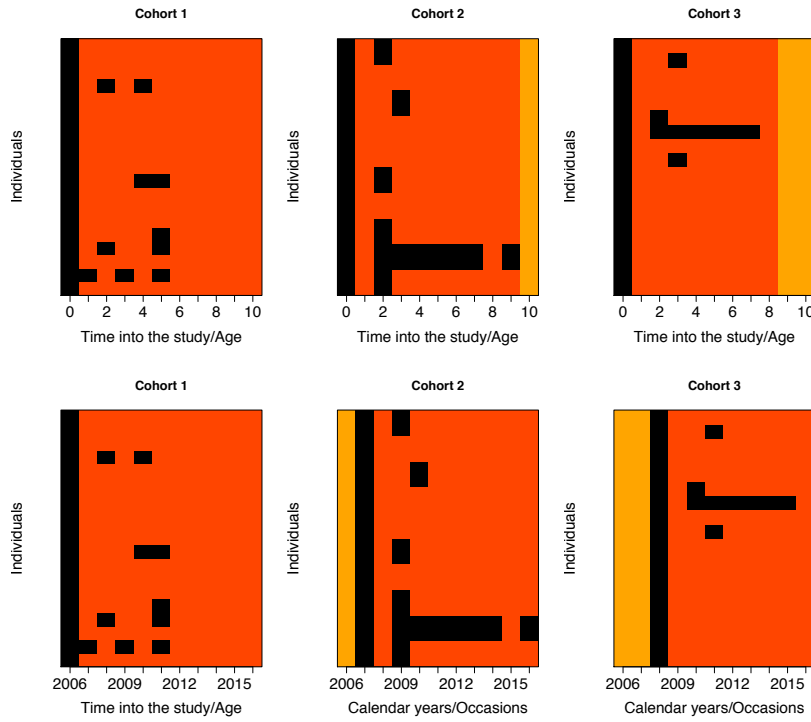


Fig. 2: Capture-histories of cohorts 1, 2 and 3 in age (top) and calendar scale (bottom) of individuals of the toy example in Figure 1. Rows represent individuals, orange lines represent non information, red lines indicate that individual has not been seen, and black lines represent that the individual has been seen. In this toy example we assume the initial age is zero years old for all the individuals (i.e.  $a_{j0} = 0$ ).

typically assumed to be dependent on the capture occasion. In particular the model parameters (allowing for additional individual heterogeneity) are given by:

$$\phi_{i,t} = \mathbb{P}(\text{individual } i \text{ is alive at time } t+1 \mid \text{alive at time } t); \text{ and} \\ p_{i,t+1} = \mathbb{P}(\text{individual } i \text{ observed at time } t+1 \mid \text{alive at time } t+1),$$

for  $i = 1, \dots, N$  and  $t = 1, \dots, T-1$ . Survival from one capture occasion to the next may change temporally due to, for example, weather conditions, food availability etc. Similarly, the capture probability may be dependent on the effort at the given occasion, or local conditions for observing individuals. For a more detailed explanation of the CJS models in calendar scale, see for example, (31), (3), (16), (27), (33) and (47), among many others.

Alternatively, the age (or time in study) scale suggests the analogous parameters given by:

$$\psi_{i,a} = \mathbb{P}(\text{individual } i \text{ is alive at age } a+1 \mid \text{alive at age } a); \text{ and} \\ \eta_{i,a+1} = \mathbb{P}(\text{individual } i \text{ observed at age } a+1 \mid \text{alive at age } a+1),$$

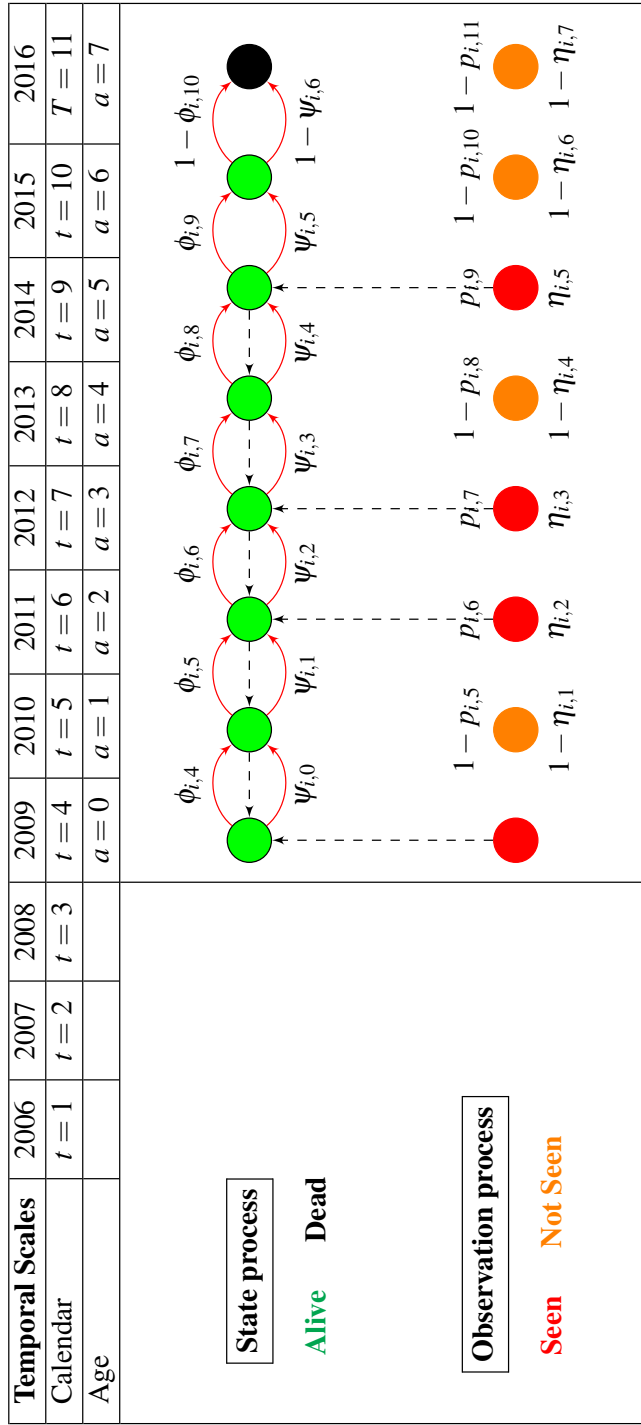
for  $i = 1, \dots, N$  and  $a = 0, \dots, A-1$ . The parameters typically reflect the life cycle of an individual, for example, from juvenile to breeding adult where the different life stages (i.e. ages) may affect both their survival and capture probabilities based on age-related behaviour. Age has been incorporated into capture-recapture analyses, with the first-year survival probability often a key component for biologically realistic models (14; 7).

The differences between both scales can be easily visualised in Table 2, where we present a particular example of an individual marked as a chick in 2009 (i.e.,  $a_{i0} = 0$ ), observed in years 2011, 2012 and 2014 (red = observed, orange = not observed), that dies in 2016 (green = alive, black = dead). Table 2 includes both scales and the associated model parameters along with the different possible values for  $t$  and  $a$ . As mentioned above, the age scale provides a more intuitive interpretation of the parameters when age is the main focus, as they are directly indicating the age of the individual: the probability that an individual aged  $a$  survives one year (instead of the probability that an individual in year  $t$  survives until the next year) and the probability that an alive individual aged  $a$  is seen (instead of the probability that an alive individual is seen on occasion  $t$ ).

Clearly parameterising the model parameters in terms of the different temporal scales leads to very different interpretations (as it can be appreciated at Table 2). Both forms of dependence may be appropriate dependent on the system; and in particular both temporal scales may be important. We describe how we can incorporate both temporal scales by extending the dependence structure of the parameters (and also consider the implications for the data representation).

### 3 CJS models with year and age temporal scales

We present the general formulation of the CJS models with combined year and age as the reference temporal scales. The year-only or age-only models can then be seen as special cases of this general case. We present two (equivalent) formulations: state-space and multinomial. The state-space formulation provides an intuitive model formulation (42; 22); whereas the multinomial formulation is mathematically efficient and permits additional absolute goodness-of-fit tests to be applied (33).

Table 2: Annual survival and resighting probabilities, calendar years  $t$  and  $a$  values for an individual ringed as chick in 2009 in both temporal scales.



### State-Space formulation

Assuming that capture occasions are annual, let  $Z_{i,a,t}$  be the latent variable that describes the true state of individual  $i$  at age  $a$  and time  $t = f_i + a - a_{i0}$  for  $a = a_{i0}, \dots, a_{i0} + T - f_i$  for all  $i = 1, \dots, N$ . We note that given the initial capture time ( $f_i$ ), associated age at initial capture ( $a_{i0}$ ) and subsequent age of the individual ( $a$ ), the capture occasion ( $t$ ) is deterministically calculated. However we retain the notation on both age ( $a$ ) and capture occasion ( $t$ ) for ease of comparison between the age and temporal scales. The possible states are alive and available for capture,  $Z_{i,a,t} = 1$ , and dead and not available for capture (i.e. dead or migrated from study),  $Z_{i,a,t} = 0$ .

The corresponding survival process is given by,

$$Z_{i,a+1,t+1} | Z_{i,a,t} = z_{i,a,t} \sim \text{Bernoulli}(z_{i,a,t} \psi_{i,a,t}), \quad (1)$$

for  $i = 1, \dots, N$  and  $a = a_{i0}, \dots, a_{i0} + T - f_i - 1$  such that  $t = f_i + a - a_{i0}$ , where  $\psi_{i,a,t}$  denotes the annual survival probability of individual  $i$  aged  $a$  at time  $t$  to time  $t + 1$ . When the interest is focused on the survival at different ages, this parameter has an intuitive interpretation, as it represents the probability that an individual aged  $a$  at time  $t$  survives one year.

Let  $Y_{i,a,t}$  denote a binary variable that describes whether individual  $i = 1, \dots, N$  is observed or not at age  $a$  and time  $t = f_i + a - a_{i0}$ , given it is alive and available for capture at time  $t$  for  $a = a_{i0} + 1, \dots, a_{i0} + T - f_i$ . The possible observations are seen ( $Y_{i,a,t} = 1$ ), and not seen ( $Y_{i,a,t} = 0$ ). The observation process is then given by,

$$Y_{i,a,t} | Z_{i,a,t} = z_{i,a,t} \sim \text{Bernoulli}(z_{i,a,t} \eta_{i,a,t}), \quad (2)$$

for  $i = 1, \dots, N$ , and  $a = a_{i0} + 1, \dots, a_{i0} + T - f_i$  such that  $t = f_i + a - a_{i0}$ , where  $\eta_{i,a,t}$  denotes the recapture probability of individual  $i$  at age  $a$  and time  $t$  given it is alive (i.e. available for capture).

Assuming independence between individuals, and conditional on the first capture, the associated likelihood function of the survival and observation states is equal to the product over each individual of the corresponding survival and observation likelihood components:

$$\begin{aligned} \ell(Y, Z | \theta) = & \prod_{i=1}^N \left[ \underbrace{\left( \prod_{a=a_{i0}}^{a_{i0}+T-f_i-1} p_z(Z_{i,a+1,f_i+a-a_{i0}+1} | Z_{i,a,f_i+a-a_{i0}}, \theta) \right)}_{\ell_{state}} \right. \\ & \left. \times \underbrace{\left( \prod_{a=a_{i0}+1}^{a_{i0}+T-f_i} p_y(Y_{i,a,f_i+a-a_{i0}} | Z_{i,a,f_i+a-a_{i0}}, \theta) \right)}_{\ell_{obs}} \right], \end{aligned} \quad (3)$$

where  $\theta = \{\psi, \eta\}$  represents the vector of all the parameters of the model,  $p_z$  and  $p_y$  are the (Bernoulli) probability functions of the state and observation processes respectively, and  $\ell_{state}$  and  $\ell_{obs}$  the likelihood terms related to the state and observation process for each individual, respectively. The observed data likelihood,  $\ell(Y | \theta)$ , is obtained by summing out over the  $Z$  values.

Alternatively in calendar scale, if we extend the dependence structure to incorporate both temporal scales (age and time), model parameters have the same interpretation as in

age scale:

$$\begin{aligned}\phi_{i,t,a} &= \mathbb{P}(\text{individual } i \text{ is alive at time } t+1 \text{ and at age } a+1 \mid \text{alive at time } t \text{ and at age } a); \\ p_{i,t+1,a+1} &= \mathbb{P}(\text{individual } i \text{ observed at time } t+1 \text{ and at age } a+1 \mid \\ &\quad \text{alive at time } t+1 \text{ and at age } a+1),\end{aligned}$$

for  $i = 1, \dots, N$  and  $t = f_i, \dots, T-1$  such that  $a = t - f_i + a_0$ , and  $\phi_{i,t,a} = \psi_{i,a,t}$ , and  $p_{i,t+1,a+1} = \eta_{i,a+1,t+1}$ .

### Multinomial formulation

Capture-recapture data are often summarised in the form of  $m$ -arrays (33). The  $m$ -array summarizes the number of individuals released at each capture occasion and subsequently next observed again at each future capture occasion, or are unobserved again within the study, and are thus typically specified in the calendar scales. When there is additional age dependence, separate  $m$ -arrays are constructed for each age, in terms of age at release. This  $m$ -array formulation is often preferred as it leads to an efficient multinomial likelihood specification, as opposed to a product over multiple Bernoulli functions, with an associated significant reduction in computing time (31). However, a particular disadvantage of this formulation is that not all models can be fitted within this framework, such as individual random effect models, as the individual capture histories are not retained in this structure. In what follows, we will explain in detail the  $m$ -arrays specified in age scale, however, for a detailed description of  $m$ -arrays in calendar scale see for example, (27; 31; 33; 47).

The  $m$ -arrays in age scale summarizes, in the form of sufficient statistics, the number of individuals released at each given age that are next captured at each subsequent age, or not observed again within the study, for each cohort within the study and age at first capture. The corresponding summary statistics correspond to an upper triangular matrix in which rows (denoted by  $a$ ) correspond to age of release and columns (denoted by  $c$ ) to age at next recapture. We note that the number of rows and columns within an  $m$ -array will depend on the cohort, as this determines how many future capture occasions are available for an individual to be released or recaptured (recall that a cohort is defined in terms of initial capture time).

We note that presenting the  $m$ -arrays in the different temporal scales an immediate difference appears: the number of  $m$ -arrays needed. In age scale, the number of  $m$ -arrays needed will differ depending on whether individuals enter into the study at different ages (general scenario) or they have a common initial age (simplest case), independently of the model used. In particular, the number of  $m$ -arrays *per cohort* is equal to the number of different initial ages observed. Alternatively, specifying the  $m$ -arrays within the calendar scale, only one  $m$ -array per age is required. Nevertheless, we note that for the age-scale in the simplest scenario where all the individuals have a common initial age, the number of  $m$ -arrays needed and their construction is arguably simpler, with only one  $m$ -array required for each separate cohort. Further, the number of rows and columns of each  $m$ -array in age scale also depends on the given associated cohort. In particular, for cohort  $w = 1, \dots, T-1$ , and initial age at capture,  $a_0$ , the rows are labelled by age,  $a$ , such that  $a = a_0, \dots, T-w+a_0$ , and the columns labelled by age at next recapture,  $c$ , such that  $c = a_0+1, \dots, T-w+a_0+1$ . Hence, the later the cohort the smaller the number of rows and columns of the associated  $m$ -array. As usual, each  $m$ -array contains an additional column corresponding to individuals never recaptured (column number  $T-w+a_0+1$  in age scale).

In age scale, the element  $(a, c)$  of the age  $m$ -array of cohort  $w$  and initial age  $a_0$ , denoted  $m_{a,c,w,a_0}$ , represents the number of individuals with initial age  $a_0$  that belong to cohort  $w$  and which were released at age  $a$  and next captured at age  $c$ , for  $a = a_0, \dots, T - w + a_0 - 1$  and  $c = a_0, \dots, T - w + a_0$ . By construction,  $m_{a,c,w,a_0} = 0$  for  $c \leq a$  (the matrix is upper triangular). Further,  $m_{a,(T-w+a_0+1),w,a_0}$  denotes the number of individuals that enter the study at initial age  $a_0$  that belong to cohort  $w$  that were released at age  $a$  and subsequently not observed again during the study (final column of the  $m$ -array). Further  $R_{a,w,a_0} = \sum_c m_{a,c,w,a_0}$  represents the number of individuals with initial age  $a_0$  that belong to cohort  $w$ , which were released at age  $a$  (i.e.  $R_{a,w,a_0}$  corresponds to the sum of the elements in row  $a$  from cohort  $w$  and initial age  $a_0$ ). Clearly,

$$m_{a,(T-w+a_0+1),w,a_0} = R_{a,w,a_0} - \sum_{c=a_0+1}^{T-w+a_0} m_{a,c,w,a_0},$$

for  $a = a_0, \dots, T - w + a_0 - 1$ . Denoting  $q_{a,c,w,a_0}$  the corresponding cell probabilities of the  $m$ -arrays, the expected values of the entries of the  $m$ -array (equal to  $R_{a,w,a_0} \times q_{a,c,w,a_0}$ ) are a function of the model parameters and the observed number of released individuals ( $R_{a,w,a_0}$ ). Finally, we note that necessarily, for each cohort,  $w$  and initial age of release  $a_0$ , the row sums of the given cell probabilities sum to unity, i.e.  $\sum_{c=a_0+1}^{T-w+a_0+1} q_{a,c,w,a_0} = 1$ , for each  $a = a_0, \dots, T - w + a_0 - 1$ . Alternatively, and for similarity with standard notation within the statistical ecology literature, we let  $\chi_{a,w,a_0}$  ( $= q_{a,T-w+a_0+1,w,a_0}$ ) denote the probability that an individual from cohort  $w$ , aged  $a_0$  at initial capture is not observed again following their release at age  $a$ .

By contrast for calendar scale the rows and the columns of the  $m$ -array (for a given age) correspond to release occasions (denoted by  $j = 1, \dots, T - 1$ ) and next subsequent recapture occasion (denoted by  $t = 2, \dots, T$ ), respectively. As before, there is an additional column, ( $j = T + 1$ ), corresponding to individuals not observed again within the study following their final release. The cell entries are typically denoted by  $m_{j,t}$ , for  $j = 1, \dots, T - 1$  and  $t = 2, \dots, T + 1$ ; with associated probabilities  $q_{j,t}$ , where once again  $\chi_j = q_{j,T+1}$ . If additional age dependence is included, we require an  $m$ -array for each age  $a$ , but note that the number of rows and columns in each corresponding  $m$ -array remain constant; notationally, this may be represented by an additional subscript for age,  $a$ , added to the associated cell entries/probabilities.

In order to clarify these concepts, Tables 3 and 4 present the  $m$ -arrays corresponding to individuals with initial age  $a_0 = 0$  of cohorts 1 and 2, respectively, in a toy example where individuals can have different initial ages, and the number of capture occasions is  $T = 6$ . For comparison, Tables 5 and 6 present  $m$ -arrays corresponding to individuals from cohort 1 in an example where all the individuals have a common initial age of 1 ( $a_0 = 1$ ) or 2 ( $a_0 = 2$ ) years, respectively, once again with  $T = 6$ . Finally, in Table 7 we present the corresponding  $m$ -array in calendar scale assuming no age dependence.

The multinomial cell probabilities differ depending on the model considered. Here we present the likelihood related to the CJS model with age and time dependent model parameters. Conditional on the numbers released and assuming independence between individuals the likelihood is,

$$\ell(Y|\theta) \propto \prod_{w=1}^{T-1} \prod_{a_0=0}^{A_0} \prod_{a=a_0}^{T-w+a_0-1} \left\{ \prod_{c=a+1}^{T-w+a_0} (q_{a,c,w,a_0})^{m_{a,c,w,a_0}} \right\} (\chi_{a,w,a_0})^{m_{a,T-w+a_0+1,w,a_0}}, \quad (4)$$

where, for  $a = a_0, \dots, T - w + a_0 - 1$  and  $c = a + 1, \dots, T - w + a_0$ ,

$$q_{a,c,w,a_0} = \prod_{k=a}^{c-1} \psi_{k,k+w-a_0} \times \prod_{k=a}^{c-2} [(1 - \eta_{k+1,k+w-a+1})] \eta_{c,c+w-a},$$

and for  $c \leq a$ ,  $q_{a,c,w,a_0} = 0$ . We note that we specify the product over the null set to be equal to 1, i.e.  $\prod_{k=a}^{a-1} \equiv 1$ ; and that the calendar time (corresponding to the second subscript of the  $\psi$  and  $\eta$  parameters) is expressed as a function of the initial age ( $a_0$ ), the age of release ( $a$ ) and cohort ( $w$ ). The first product in the above expression for  $q_{a,c,w,a_0}$  corresponds to an individual surviving from age  $a$  to age  $c$ ; the second product to not being observed between times  $a$  and  $c - 1$ , before being subsequently observed at time  $c$ . For completeness we note that the probability an individual from cohort  $w$ , with initial age  $a_0$  at age that is released at age  $a$  is not observed again within the study can be expressed as,

$$\chi_{a,w,a_0} = 1 - \sum_{c=a+1}^{T-w+a_0} (q_{a,c,w,a_0}).$$

Finally we note that, further restrictions may be specified on these parameters to represent age classes, rather than distinct ages.

Therefore, in the capture-recapture framework this model can be fitted in both calendar and/or age scales. However, working with this model and under multinomial formulation, the differences between scales are more remarkable regarding the presentation of the data on the [age](#) or time scales (i.e. number of  $m$ -arrays needed).

We note that when the (full) age and/or time dependence is included in both the recapture and survival probabilities, the corresponding CJS model is intrinsically parameter redundant (33; 15). In particular, two parameters are confounded and not uniquely estimable, and only their product is estimable. The non identifiable parameters are  $\phi_{T-1}$  and  $p_T$  (in calendar scale), and  $\psi_{A-1}$  and  $\eta_A$  (in age scale, assuming the maximum age,  $A$ , is a single age, and not a set of ages). Further identifiability issues can arise due to the observed data (when 0 cell entries are observed), for further discussion of these issues see for example, (10).

#### 4 Survival analysis of a colony of common guillemot

In order to illustrate the differences and similarities between the age and calendar temporal scales, we consider two models applied to a real data set. In particular we focus on (i) the year-only and age-only models in calendar and age scale, respectively, and (ii) the time and age dependent model (equivalent for both the age and calendar scales). We present the analysis of a database obtained from monitoring programs carried out by the Baltic Seabird Project from 2006 until 2016 (so that  $T = 11$ ) in the largest colony of common guillemot in the Baltic Sea (Stora Karlsö, Sweden). For each year of the study chicks were captured and individually ringed (with metallic and plastic rings) after jumping from the ledges at an age of about 20 days. Thus, for this study all the individuals were ringed at a common initial age of zero years old, i.e.  $a_{i0} = 0$  for  $i = 1, \dots, N$ . The total number of individuals ringed during this study period was  $N = 28\,930$ . During their reproductive period (May to July), resightings with telescopes and binoculars were made from above the breeding ledges with minimal disturbance in order to identify marked individuals through the reading of the metallic and/or plastic ring codes (recaptures). We note that for this example we work with the simplest scenario where we need only construct one  $m$ -array per cohort (a total

Table 4:  $M$ -array for initial age  $a_0 = 0$  and cohort 2 for an study where individuals can have different initial ages. Released individuals ( $R_{a,w,a_0}$ ), observed individuals ( $m_{a,c,w,a_0}$ ) and cell probabilities ( $q_{a,c,w,a_0}$  and  $\chi_{a,w,a_0}$ ).

Cohort 2 ( $a_0 = 0$ )							
Age at first recapture							
	1	2	3	4	5	6	Never seen
0	$R_{0,2,0}$	$m_{0,1,2,0}$	$m_{0,2,2,0}$	$m_{0,3,2,0}$	$m_{0,4,2,0}$	—	$m_{0,5,2,0}$
		$(q_{0,1,2,0})$	$(q_{0,2,2,0})$	$(q_{0,3,2,0})$	$(q_{0,4,2,0})$	—	$(\chi_{0,2,0})$
1	$R_{1,2,0}$	—	$m_{1,2,2,0}$	$m_{1,3,2,0}$	$m_{1,4,2,0}$	—	$m_{1,5,2,0}$
			$(q_{1,2,2,0})$	$(q_{1,3,2,0})$	$(q_{1,4,2,0})$	—	$(\chi_{1,2,0})$
$\vdots$			$\vdots$	$\vdots$	$\vdots$		$\vdots$
3	$R_{3,2,0}$	—	—	—	$m_{3,4,2,0}$	—	$m_{3,5,2,0}$
					$(q_{3,4,2,0})$	—	$(\chi_{3,2,0})$

Table 6:  $M$ -array for initial age  $a_0 = 2$  and cohort 1 in an study where all the individuals have a common initial age  $a_0 = 2$ . Released individuals ( $R_{a,w,a_0}$ ), observed individuals ( $m_{a,c,w,a_0}$ ) and cell probabilities ( $q_{a,c,w,a_0}$  and  $\chi_{a,w,a_0}$ ).

Cohort 1 (common $a_0 = 2$ )							
Age at first recapture							
	3	4	5	6	7	8	Never seen
2	$R_{2,1,2}$	$m_{2,3,1,2}$	$m_{2,4,1,2}$	$m_{2,5,1,2}$	$m_{2,6,1,2}$	$m_{2,7,1,2}$	$m_{2,8,1,2}$
		$(q_{2,3,1,2})$	$(q_{2,4,1,2})$	$(q_{2,5,1,2})$	$(q_{2,6,1,2})$	$(q_{2,7,1,2})$	$(\chi_{2,1,2})$
3	$R_{3,1,2}$	—	$m_{3,4,1,2}$	$m_{3,5,1,2}$	$m_{3,6,1,2}$	$m_{3,7,1,2}$	$m_{3,8,1,2}$
			$(q_{3,4,1,2})$	$(q_{3,5,1,2})$	$(q_{3,6,1,2})$	$(q_{3,7,1,2})$	$(\chi_{3,1,2})$
$\vdots$			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
6	$R_{6,1,2}$	—	—	$m_{6,5,1,2}$	$m_{6,6,1,2}$	$m_{6,7,1,2}$	$m_{6,8,1,2}$
				$(q_{6,5,1,2})$	$(q_{6,6,1,2})$	$(q_{6,7,1,2})$	$(\chi_{6,1,2})$

Table 3:  $M$ -array for initial age  $a_0 = 0$  and cohort 1 for an study where individuals can have different initial ages. Released individuals ( $R_{a,w,a_0}$ ), observed individuals ( $m_{a,c,w,a_0}$ ) and cell probabilities ( $q_{a,c,w,a_0}$  and  $\chi_{a,w,a_0}$ ).

Cohort 1 ( $a_0 = 0$ )							
Age at first recapture							
	1	2	3	4	5	6	Never seen
0	$R_{0,1,0}$	$m_{0,1,1,0}$	$m_{0,2,1,0}$	$m_{0,3,1,0}$	$m_{0,4,1,0}$	$m_{0,5,1,0}$	$m_{0,6,1,0}$
		$(q_{0,1,1,0})$	$(q_{0,2,1,0})$	$(q_{0,3,1,0})$	$(q_{0,4,1,0})$	$(q_{0,5,1,0})$	$(\chi_{0,1,0})$
1	$R_{1,1,0}$	—	$m_{1,2,1,0}$	$m_{1,3,1,0}$	$m_{1,4,1,0}$	$m_{1,5,1,0}$	$m_{1,6,1,0}$
			$(q_{1,2,1,0})$	$(q_{1,3,1,0})$	$(q_{1,4,1,0})$	$(q_{1,5,1,0})$	$(\chi_{1,1,0})$
$\vdots$			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
4	$R_{4,1,0}$	—	—	—	$m_{4,5,1,0}$	$m_{4,6,1,0}$	$m_{4,7,1,0}$
					$(q_{4,5,1,0})$	$(q_{4,6,1,0})$	$(\chi_{4,1,0})$

Table 5:  $M$ -array for initial age  $a_0 = 1$  and cohort 1 in an study where all the individuals have a common initial age  $a_0 = 1$ . Released individuals ( $R_{a,w,a_0}$ ), observed individuals ( $m_{a,c,w,a_0}$ ) and cell probabilities ( $q_{a,c,w,a_0}$  and  $\chi_{a,w,a_0}$ ).

Cohort 1 (common $a_0 = 1$ )							
Age at first recapture							
	2	3	4	5	6	7	Never seen
1	$R_{1,1,1}$	$m_{1,2,1,1}$	$m_{1,3,1,1}$	$m_{1,4,1,1}$	$m_{1,5,1,1}$	$m_{1,6,1,1}$	$m_{1,7,1,1}$
		$(q_{1,2,1,1})$	$(q_{1,3,1,1})$	$(q_{1,4,1,1})$	$(q_{1,5,1,1})$	$(q_{1,6,1,1})$	$(\chi_{1,1,1})$
2	$R_{2,1,1}$	—	$m_{2,3,1,1}$	$m_{2,4,1,1}$	$m_{2,5,1,1}$	$m_{2,6,1,1}$	$m_{2,7,1,1}$
			$(q_{2,3,1,1})$	$(q_{2,4,1,1})$	$(q_{2,5,1,1})$	$(q_{2,6,1,1})$	$(\chi_{2,1,1})$
$\vdots$			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
5	$R_{5,1,1}$	—	—	—	$m_{5,6,1,1}$	$m_{5,7,1,1}$	$m_{5,8,1,1}$
					$(q_{5,6,1,1})$	$(q_{5,7,1,1})$	$(\chi_{5,1,1})$

Table 7: General  $m$ -array in calendar scale: released ( $R_j$ ), observed individuals ( $m_{j,t}$ ) and cell probabilities ( $q_{j,t}$  and  $\chi_j$ ), for  $T = 6$ 

		First Recapture occasion						
		2	3	4	5	6	Never seen	
Release occ.	1	$R_1$	$m_{1,2}$ $(q_{1,2})$	$m_{1,3}$ $(q_{1,3})$	$m_{1,4}$ $(q_{1,4})$	$m_{1,5}$ $(q_{1,5})$	$m_{1,6}$ $(q_{1,6})$	$m_{1,7}$ $(\chi_1)$
	2	$R_2$	—	$m_{2,3}$ $(q_{2,3})$	$m_{2,4}$ $(q_{2,4})$	$m_{2,5}$ $(q_{2,5})$	$m_{2,6}$ $(q_{2,6})$	$m_{2,7}$ $(\chi_2)$
	$\vdots$			$\vdots$				$\vdots$
	5	$R_5$	—	—	—	—	$m_{3,6}$ $(q_{3,6})$	$m_{3,7}$ $(\chi_5)$

of  $T - 1 = 10$ ) for the age scale representation; for the calendar scale we construct one  $m$ -array for each possible age a bird is released within the study (which without age groups corresponds to a maximum of  $T - 1 = 10$  ages).

In order to highlight differences between temporal scales, we will not establish age classes in the age-only or time-only CJS models. Nevertheless, we do establish age classes in the CJS model where both temporal scales are considered. We let  $a = 0, 1, 2, 3+$  denote the age classes, being those related to individuals in their first year of life, second, third and finally, individuals in their fourth or more years of life (adults), respectively. To illustrate the differences in data presentation when using the different temporal scales we present the cohort  $m$ -arrays and the age  $m$ -arrays related to age and calendar scales (for conciseness, we present the calendar scale for the age classes given above), in Appendix A.

The set of parameters will differ for each model proposed and the temporal scale considered (for the age-only or time-only models). The model parameters of the full time (or age) dependent CJS models are,  $\psi = \{\psi_0, \psi_1, \dots, \psi_9\}$  and  $\eta = \{\eta_1, \eta_2, \dots, \eta_{10}\}$ , in age scale; and  $\phi = \{\phi_1, \phi_2, \dots, \phi_{10}\}$  and  $\mathbf{p} = \{p_2, p_3, \dots, p_{11}\}$ , in calendar time scale. However, parameters  $\psi_9$  and  $\eta_{10}$  in age scale; and similarly,  $\phi_{10}$  and  $p_{11}$  in calendar scale, are **confounded** (i.e. we can estimate the product of the corresponding parameters but not each one separately). For the age and time dependent model we specify additive age and time effects (on the logit scale) for both survival and recapture probabilities such that  $\text{logit}(\psi_{a,t}) = \alpha_a + \beta_t$ , and  $\text{logit}(\eta_{a+1,t+1}) = \alpha'_a + \beta'_t$ , for  $a = 0, \dots, 3+$  and  $t = 1, \dots, T - 1$ . Thus, the corresponding set of model parameters for the CJS model with age and time dependence are,  $\alpha = \{\alpha_0, \dots, \alpha_{3+}\}$  and  $\beta = \{\beta_1, \dots, \beta_{10}\}$  (for the survival probabilities) and  $\alpha' = \{\alpha'_0, \dots, \alpha'_{3+}\}$  and  $\beta' = \{\beta'_1, \dots, \beta'_{10}\}$  (for the capture probabilities). Finally, for identifiability, we set  $\alpha_0 = \alpha'_0 = 0$ , and subsequently note that  $\alpha_a$  and  $\alpha'_a$ , for  $a = 1, 2, 3+$  are interpreted as the effect for age  $a$ , relative to age 0; and that the corresponding  $\beta_t$  and  $\beta'_t$  terms are thus interpreted as the associated (baseline) time effects for an individual of age 0.

## Bayesian inference

Maximum likelihood estimation or Bayesian methods can be used to fit the CJS models to the capture-recapture data to obtain inference on the model parameters (27; 36; 33; 47,

and references therein). In our case, we apply a Bayesian approach, as this permits us to make probabilistic statements about the model parameters and, more generally, provides the mechanism for explicitly incorporating available prior knowledge about the unknown parameters into the inferential process; see for example, (32; 27; 9; 17).

The resulting posterior distribution of the parameters, given the data, contains all the information about the system but is not available in closed form. Thus we apply a Markov chain Monte Carlo (MCMC) approach (41) to learn about the posterior distribution. For our application, we implemented the MCMC algorithm via the `jags` software (38) within the R program (40). As no prior knowledge about the parameters was available, we specified non informative prior distributions for all the parameters of the model. In particular, we specify Uniform(0,1) prior distributions for all survival and recapture probabilities in the CJS models with only-time or only-age dependence; and Normal(0,10) prior distributions for the age ( $\alpha_a$  and  $\alpha'_a$ ) and time ( $\beta_t$  and  $\beta'_t$ ) effect parameters in the CJS model with age and time dependence. To avoid identifiability issues, we set  $\alpha_0 = \alpha'_0 = 0$ . Finally, initial values are specified for the model parameters, these are simulated from Uniform(0,1) distributions for the recapture probabilities and Normal(0,10) distributions for the  $\alpha_a$ ,  $\alpha'_a$ ,  $\beta_t$  and  $\beta'_t$  parameters. Three independent chains are run, each for 200,000 iterations, discarding the first 20,000 as burn-in (to ensure that the Markov chain has reached the stationary distribution) and thinning every 100 of the iterations for the memory storage purposes and to reduce auto correlation. The R code is available from the author on request.

## Results

Figure 3 provides a graphical representation of posterior distribution of survival (cyan) and recapture (blue) probabilities for both temporal scales considering full time dependence (on age or time). Due to the identifiability problem mentioned, in this Figure we only present the estimable parameters. As can be seen from Figure 3 (and highlighted in Table 2), if we only incorporate time or age dependence the model parameters differ significantly between the different temporal scales. In particular, for the age scale model parameters represent differences in survival and recapture probabilities corresponding to the age of individuals, whereas conversely, in calendar scale the parameters represent interannual variations in survival and resighting probabilities.

If we extend the dependence to age and time, the corresponding model parameters are equivalent in both temporal scales, as there are both age and time components. Figure 4 provides the corresponding estimated posterior means and associated 95% credible intervals for the survival probabilities (on the left) and recapture probabilities (on the right) for each year in the study and possible age class. We note that, unsurprisingly, the lowest survival probabilities correspond to individuals in their first year of life (i.e. age  $a = 0$ ). These results agree with previous studies that show that first year survival is lower than older individuals (18), and that survival probabilities typically vary inter-annually due to variation in food availability and environmental conditions (48; 19). Our results are in line with these studies, showing a marked variation between years. Similarly, age is commonly a very important factor affecting survival and recapture probabilities (see for example, 12; 18; 35). However, it is noticeable the (unrealistic) high estimates obtained for the survival probability at age 2 as well as the (unexpected) high recapture probabilities and low survival probabilities estimated for ages 3+. Further investigation suggests that the high survival probability at age 2 is an consequence and artefact of the low number of individuals observed age 1 (i.e. the year following their initial capture). We note that the lowest recapture probability occurs

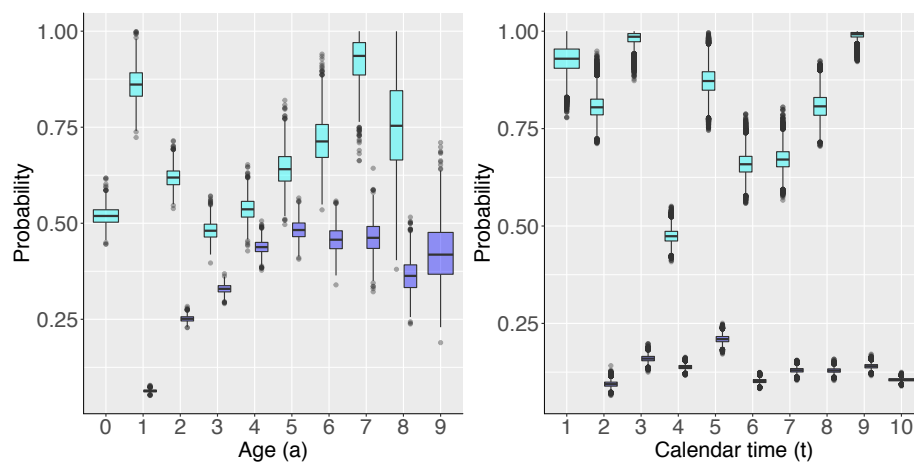


Fig. 3: Graphical representation of the marginal posterior distributions for the annual survival (cyan) and resighting probabilities (blue) probabilities in the full time dependent CJS model in age scale (left) and calendar time scale (right). The final box (for age 9 or time 10), corresponds to the product of the associated confounded final survival and recapture probabilities.

at age 1 (see Figures 4 and 3), due to the relatively small number of birds recaptured at this time. This aspect is common for this species, which, after fledging, immature birds spend several years at sea until they can back to the colony (13; 12). However, in this colony a large number of individuals are subsequently observed aged 2. As a consequence, we estimate a reduced first-year survival probability (and hence expected number of birds at age 1), which to then account for the large number of birds observed at age 2 leads to the very large (and unrealistic) second year survival probabilities. Finally, the recapture and the survival probabilities for individuals aged 3+ (i.e. adults) generally appear to be an overestimate and underestimate, respectively. The primary reason for these issues appears to be a result of the capture-recapture protocol at the colony-level in terms of the partial monitoring system implemented for recapturing individuals. For further discussion of this issue (and others), and associated additional modelling approaches to address these issues see (43; 44).

## 5 Conclusions

Measuring time to the occurrence of events, typical for survival analysis, is often an area of particular interest. In this paper we connect two distinct areas in the context of survival analysis: human demography and population ecology. For these different areas, there is often a different emphasis given to the elicitation of the temporal scale. In medical studies the origin time and consequently, the temporal scale, must be defined at the beginning of the study. In this context, there are several time scales defined in relation to the possible times zero (20). However, in ecology these procedures are not performed, and consequently, the default is the capture-occasion (or calendar) scale. In this work, we adopt the approach and data presentation used in medical studies and apply this to the ecological capture-recapture framework. By doing so we present the individual capture-histories in an alternative temporal scale, corresponding to age (as opposed to capture occasion). Typically, age is an



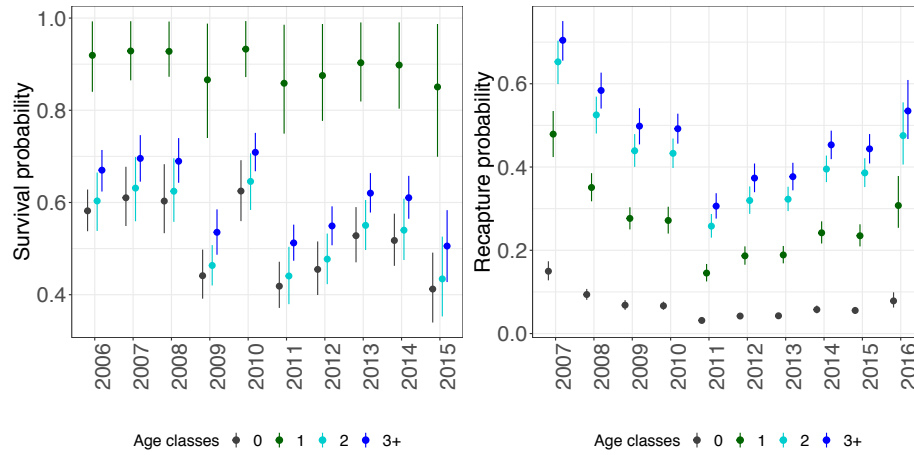


Fig. 4: Mean and 95% symmetric credible interval of the marginal posterior distributions of the annual survival probabilities (left) and recapture probabilities (right) for each year of the study for each age class for the combined age and time CJS model.

important factor affecting survival probabilities in animal populations and consequently, the survival and recapture probabilities may be expressed as a function of age, in addition to time. Additionally, in many capture-recapture protocols (as for the real example presented in this work) individuals are marked days after they are born (as it may be easier to capture such individuals from their breeding sites), with time zero then set equal to birth time. From medical framework perspective, if time zero is birth time the ‘natural’ temporal scale is then simply age.

In ecology, different models can be used to estimate survival and recapture probabilities from capture-recapture data. Although it can be easily extended to other models, we consider the important Cormack-Jolly-Seber (CJS) model. We present the general notation of the CJS model and the associated likelihood function in terms of both the commonly used state-space and multinomial formulations using age as the temporal scale. We consider a real database corresponding to a large colony of common guillemots in the Baltic Sea and fit the data to: (i) the CJS model with full time dependence (being age or time, depending on the temporal scale used), and (ii) the CJS model with the model parameters both age and time dependent.

The natural presentation of the observed capture histories can be regarded as dependent on the temporal scale considered. For the calendar scale, the capture histories are typically specified to be of length  $T$  (i.e. the number of capture occasions) for all individuals; whereas for the age scale, the trailing zeroes may be omitted in which case the capture histories will be of different lengths (where the length is dependent on the cohort). In addition the age at initial capture is also required to complete the data specification. Alternatively, if we present the data in terms of the  $m$ -array summary statistics, further differences can be noted. One of the interesting differences between temporal scales is the number of  $m$ -arrays needed in the multinomial formulation. In particular in calendar scale, when no age dependence is present in any of the parameters of the model, we only need to construct one general  $m$ -array to summarize the whole data set; alternatively, if age-dependence is present, we need a  $m$ -array for each age class. However, when presenting the  $m$ -arrays using the age scale, the number of  $m$ -arrays depends on the number of different initial ages observed for the individuals,

irrespective of whether or not the parameters are age dependent. If the individuals have different ages at the beginning of the study, this [temporal](#) age scale leads to one  $m$ -array for each initial age per cohort; thus if all the individuals have a common initial age, only one  $m$ -array per cohort is needed.

In the medical framework, it is well known that changes in temporal scales may lead to different inferences and parameter interpretations (29; 8; 49). In the ecological capture-recapture framework, when model parameters are only time or age dependent they have a different interpretation depending on the temporal scale used. The alternative temporal scale presented in this work directly takes into account the age effect in survival and resighting probabilities, providing a more understandable interpretation of the age-dependent model parameters. Indeed, [due to the data are more naturally associated with the model on the same temporal scale, we consider that the age scale should be taken into consideration in studies where age is one of the main concerns or whether inter-annual variations in survival are not considered \(e.g. in cases where the population trajectory is stable\). However, when calendar time is under consideration, either the model formulation or  \$m\$ -arrays construction in calendar time scale are simpler to perform. However, if we generalize this model, extending the temporal structure to both age and time, then both scales are incorporated within the analysis and there is no distinction between scales.](#)

To conclude, medical and ecological survival studies are often assumed very distant areas particularly in terms of temporal scales but in this work we bring these ideas together. [We present an alternative temporal representation for the ecological capture-recapture framework. The methodology presented in this paper can be easily extended to other models within the capture-recapture framework. In particular, when age is the main interest in the study, this alternative scale may provide a better insight into the data representation in addition to the model parameters. Further comparisons between the application of survival analysis ideas across the medical and ecological application areas are likely to provide further advances. An example of this can be seen, for example, with the application of the ideas associated with competing risk models in medical studies, applied to spatially explicit capture-recapture in terms of which trap an individual is observed by given an individual is observed \(2\).](#)

**Acknowledgements** Blanca Sarzo has a research grant (BES-2014-070766) supported by MTM2013-42323-P from the Spanish Ministry of Economy and Competitiveness. This work has been partially supported by grants MTM2016-77501-P and TEC2016-81900-REDT from the Spanish Ministerio de Ciencia, Innovación y Universidades. Agencia Estatal de Investigación (jointly financed by the European Regional Development Fund, FEDER). RK was funded by a Leverhulme Research Fellowship. Field work on Stora Karlsö has been made possible through a long-term engagement in the Baltic Seabird project by WWF Sweden. We thank a large number of field workers and volunteers at Stora Karlsö in the period 2006-2016 for support with ringing and observation studies. Karlsö Jagt-och Djurskyddsfrenings AB provided logistical support.

## Conflict of interest

The authors declare that they have no conflict of interest.

## References

1. Bonner, S.J., Schwarz, C.J.: An extension of the CormackJollySeber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics* **62**, 142–149

2. Borchers, D.L., Efford, M.G.: Spatially explicit maximum likelihood methods for capture-recapture studies. *Biometrics* **64**, 377–85 (2008)
3. Brooks, S.P., Catchpole, E.A., Morgan, B.J.: Bayesian animal survival estimation. *Statistical Science* **15**(4), 357–376 (2000)
4. Brownie, C., Hines, J.E., Nichols, J.D., Pollock, K.H., Hestbeck, J.B.: Capture-recapture studies for multiple strata including non-Markovian transitions. *Biometrics* **49**, 1173–1187 (1993)
5. Burnham, K.: A theory for combined analysis of ring recovery and recapture data. In: J. Lebreton, P.M. North (eds.) *Marked Individuals in the Study of Bird Population*, pp. 199–213. Birkhuser Verlag Basel, Switzerland (1993)
6. Catchpole, E., Freeman, S.N., Morgan, B.J.T., Harris, M.P.: Integrated recovery/recapture data analysis. *Biometrics* **54**, 33–46 (1998)
7. Catchpole, E.A., Morgan, B.J.T.: Model selection in ring-recovery models using score tests. *Biometrics* **52**(2), 664–672 (1996)
8. Cheung, Y.B., Gao, F., Khoo, K.S.: Age at diagnosis and the choice of survival analysis methods in cancer epidemiology. *Journal of Clinical Epidemiology* **56**(1), 38–43 (2003). DOI 10.1016/S0895-4356(02)00536-X
9. Christensen, R., Johnson, W., Branscum, A., Hanson, T.E.: *Bayesian ideas and Data analysis: An introduction for scientists and statisticians*. CRC Press (2011)
10. Cole, D.J., Morgan, B.J., Titterton, D.: Determining the parametric structure of models. *Mathematical biosciences* **228**(1), 16–30 (2010)
11. Cormack, R.: Estimates of survival from the sighting of marked animals. *Biometrika* **51**, 429–438 (1964)
12. Crespín, L., Harris, M.P., Lebreton, J.D., Frederiksen, M., Wanless, S.: Recruitment to a seabird population depends on environmental factors and on population size. *Journal of Animal Ecology* **75**(1), 228–238 (2006)
13. Croxall, J., Rothery, P.: Population regulation of seabirds: implications of their demography for conservation. *Bird population studies: relevance to conservation and management*. Oxford University Press, Oxford pp. 272–296 (1991)
14. Freeman, S.N., Morgan, B.J.T.: A modelling strategy for recovery data from birds ringed as nestlings. *Biometrics* **48**(1), 217–235 (1992)
15. Giménez, O., Morgan, B.J., Brooks, S.P.: Weak identifiability in models for mark-recapture-recovery data. In: *Modeling demographic processes in marked populations*, pp. 1055–1067. Springer (2009)
16. Giménez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H., Vila, J.P., Pradel, R.: State-space modelling of data on marked individuals. *Ecological Modelling* **206**(3), 431–438 (2007)
17. Golchi, S.: Informative priors and Bayesian computation. In: *Data Science and Advanced Analytics (DSAA), 2016 IEEE International Conference*, pp. 782–789 (2016)
18. Harris, M.P., Frederiksen, M., Wanless, S.: Within-and between-year variation in the juvenile survival of Common Guillemots *Uria aalge*. *Ibis* **149**(3), 472–481 (2007)
19. Harris, M.P., Wanless, S., Rothery, P., Swann, R.L., Jardine, D.: Survival of adult common guillemots *Uria aalge* at three scottish colonies. *Bird Study* **47**(1), 1–7 (2000)
20. Hougaard, P.: *Analysis of multivariate survival data*. Springer (2000)
21. Jolly, G.: Explicit estimates from capture-recapture data with both death and immigration–stochastic model. *Biometrika* **52**, 225–247 (1965)
22. King, R.: Statistical Ecology. *Annual Review of Statistics and Its Application* **1**(1), 401–426 (2014). DOI <http://dx.doi.org/10.1146/annurev-statistics-022513-115633>

23. King, R., Brooks, S.P.: Closed form likelihoods for Arnason-Schwarz models. *Biometrika* **90**, 435–444 (2003)
24. King, R., Brooks, S.P., Coulson, T.: Analysing Complex Capture-Recapture Data: Incorporating Time-varying Covariate Information. *Biometrics* **64**, 1187–1195 (2008)
25. King, R., Langrock, R.: Semi-Markov Arnason-Schwarz models. *Biometrics* **72**, 619–628 (2016)
26. King, R., McCrea, R.S.: A generalised likelihood framework for partially observed capture-recapture-recovery models. *Statistical Methodology* **17**, 30–45 (2014). DOI <https://doi.org/10.1016/j.stamet.2013.07.004>. URL <http://www.sciencedirect.com/science/article/pii/S1572312713000609>. Modern Statistical Methods in Ecology
27. King, R., Morgan, B., Giménez, O., Brooks, S.: Bayesian Analysis for Population Ecology. CRC Press (2010)
28. Kom, E., Graubard, B., Midthune, D.: Time-to-Event Analysis of Longitudinal Follow-up of a Survey: Choice of the Time-scale. *American Journal of Epidemiology* **145**(1), 72–80 (1997)
29. Lamarca, R., Alonso, J., Gómez, G., Muñoz, Á.: Left-truncated data with age as time scale: An alternative for survival analysis in the elderly population. *The Journals of Gerontology: Series A* **53A**(5), 337–343 (1998). DOI 10.1093/gerona/53A.5.M337. URL <http://dx.doi.org/10.1093/gerona/53A.5.M337>
30. Langrock, R., King, R.: Maximum likelihood estimation of mark-recapture-recovery models in the presence of continuous covariates. *The Annals of Applied Statistics* **7**(3), 1709–1732 (2013)
31. Lebreton, J., Burnham, K., Clobert, J., Anderson, D.: Modeling survival and testing biological hypothesis using marked animals: A unified approach with case studies. *Ecological Monographs* **62**, 67–118 (1992)
32. McCarthy, M.A., Masters, P.: Profiting from prior information in Bayesian analyses of ecological data. *Journal of Applied Ecology* **42**, 1012–1019 (2005)
33. McCrea, R., Morgan, B.: Analysis of Capture-recapture Data. CRC Press (2015)
34. Morgan, B.J.T., Freeman, S.N.: A model with first-year variation for ring-recovery data. *Biometrics* **45**, 1087–1102 (1989)
35. Morrison, K.W., Hipfner, J.M., Gjerdrum, C., Green, D.J.: Wing length and mass at fledging predict local juvenile survival and age at first return in tufted puffins. *The Condor* **111**(3), 433–441 (2009)
36. Newman, K., Buckland, S., Morgan, B., King, R., Borchers, D., Cole, D., Besbeas, P., Gimenez, O., Thomas, L.: Modelling Population Dynamics. Springer (2014)
37. Pledger, S., Efford, M., Pollock, K.H., Collazo, J.A., Lyons, J.E.: Stopover duration analysis with departure probability dependent on unknown time since arrival. In: D.L. Thomson, E.G. Cooch, M.J. Conroy (eds.) *Modeling demographic processes in marked populations*, pp. 349–363. Springer (2009)
38. Plummer, M.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: *Proceedings of the 3rd international workshop on distributed statistical computing*, vol. 124, pp. 1–9 (2003)
39. Pradel, R.: Multievent: An extension of multistate capture-recapture models to uncertain states. *Biometrics* **61**, 442–447 (2005)
40. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2020). URL <https://www.R-project.org/>

41. Robert, C., Casella, G.: A short history of Markov chain Monte Carlo: Subjective recollections from incomplete data. *Statistical Science* **26**(1), 102–115 (2011)
42. Royle, J.A.: Modeling individual effects in the Cormack-Jolly-Seber model: A state-space formulation. *Biometrics* **64**(2), 364–370 (2008)
43. Sarzo, B., Armero, C., Conesa, D., Hentati-Sundberg, J., Olsson, O.: Bayesian immature survival analysis of the largest colony of Common murre *Uria aalge* in the Baltic sea. *Waterbirds* **42**(3), 304–313 (2019)
44. Sarzo, B., King, R., Conesa, D., Hentati-Sundberg, J.: Correcting bias in survival probabilities for partially monitored populations via integrated models. Submitted (2020)
45. Schwarz, C.J., Schweigert, J.F., Arnason, A.N.: Estimating migration rates using tag-recovery data. *Biometrics* **49**, 177–193 (1993)
46. Seber, G.: A note on the multiple recapture census. *Biometrika* **52**, 249–259 (1965)
47. Seber, G., Schofield, M.: *Capture-Recapture: Parameter Estimation for Open Animal Populations*. Springer (2019)
48. Votier, S.C., Birkhead, T.R., Oro, D., Trinder, M., Grantham, M., Clark, J.A., McCleery, R.H., Hatchwell, B.J.: Recruitment and survival of immature seabirds in relation to oil spills and climate variability. *Journal of Animal Ecology* **77**(5), 974–983 (2008)
49. Westreich, D., Cole, S.R., Tien, P.C., Chmiel, J.S., Kingsley, L., Funk, M.J., Anastos, K., Jacobson, L.P.: Time scale and adjusted survival curves for marginal structural Cox models. *American Journal of Epidemiology* **171**(6), 691–700 (2010). DOI 10.1093/aje/kwp418. URL <http://dx.doi.org/10.1093/aje/kwp418>

## A Appendix

In order to highlight differences in data presentation when using both temporal scales and for reproducibility issues, here we present the ten cohort  $m$ -arrays in age scale (Tables 8 to 17) and the four age  $m$ -arrays in calendar scale (Tables 18 to 21), corresponding to the study database.

[illegible]

	Cohort 2
	Age at first recapture
	1    2    3    4    5    6    7    8    9    10    Never seen
0	50   290   175   51   18   8   5   2   4   -   1362
1	0   12   4   2   0   0   0   0   0   -   32
2	0   0   82   16   5   1   1   0   0   -   197
3	0   0   63   10   1   0   1   2   -   184
4	0   0   0   0   43   4   1   0   1   -   83
5	0   0   0   0   0   34   4   2   1   -   35
6	0   0   0   0   0   0   23   3   3   -   19
7	0   0   0   0   0   0   0   12   7   -   15
8	0   0   0   0   0   0   0   0   6   -   14
9	-   -   -   -   -   -   -   -   -   -   0

[illegible][illegible]

[illegible][illegible][illegible][illegible]



Table 17:  $M$ -array cohort 10 in age scale.

[illegible]

Table 19: Age 2  $m$ -array in calendar scale.

[illegible]

Table 21: Age 4+  $m$ -array in calendar scale.

[illegible]Table 18: Age 1  $m$ -array in calendar scale.[illegible]

Table 20: Age 3  $m$ -array in calendar scale.

First Recapture occasion												
	2	3	4	5	6	7	8	9	10	11	Never seen	
1	-	-	-	-	-	-	-	-	-	-	-	
2	-	-	-	-	-	-	-	-	-	-	-	
3	-	-	-	-	-	-	-	-	-	-	-	
4	0	0	0	96	23	6	2	0	0	2	154	
5	0	0	0	0	124	18	5	2	2	2	305	
6	0	0	0	0	0	122	20	9	2	2	287	
7	0	0	0	0	0	0	110	17	7	6	252	
8	0	0	0	0	0	0	0	139	34	15	313	
9	0	0	0	0	0	0	0	0	152	38	391	
10	0	0	0	0	0	0	0	0	0	178	436	